DETECTION AND SUMMATION OF PUS CELL FOR SPUTUM QUALITY TESTING

NORAZURA BINTI ABDUL HALIM

A thesis submitted in fulfillment

of requirements for the award of the degree of

Bachelor of Electrical Engineering (Electronics)

Faculty of Electrical and Electronics Engineering

Universiti Malaysia Pahang

JUNE 2012

ABSTRAK

Penyakit yang berkaitan dengan paru-paru seperti *Moraxella catarrhalis*, *Mycobacterium tuberculosis* dan lain-lain boleh diketahui melalui kahak. Walaubagaimanapun, sampel kahak perlu melalui kultur proses yang menelan belanja yang tinggi sebelum penyakit-penyakit di atas dapat diketahui. Maka, ujian terhadap kualiti kahak harus dijalankan untuk mengelakkan berlakunya sebarang pembaziran. Hanya kahak yang berkualiti atau positif sahaja yang akan menjalani proses ini. Projek ini dijalankan untuk menggantikan kaedah manual yang diamalkan di Hospital Universiti Sains Malaysia (HUSM) untuk menentukan kualiti kahak berdasarkan 'Bartlett Criteria'. Kaedah manual merujuk kepada penilaian kualiti sesuatu kahak dengan melihat sampel kahak tersebut melalui mikroskop. Jurumakaml akan mengira bilangan sel nanah yang terdapat di dalam sampel kahak melalui mikroskop untuk memenuhi 'Bartlett Criteria'. Maka, satu system berdasarkan pemprosesan imej yang mampu untuk mengesan dan mengira bilangan sel nanah secara automatik dibangunkan untuk menggantikan kaedah manual ini. Sistem ini memerlukan empat imej kahak yang mewakili satu sampel kahak. Kesemua imej ini akan diproses satu per satu melalui sistem ini. Sistem ini juga akan mengira bilangan sel nanah yang terdapat di dalam setiap imej kahak. Akhirnya, sistem ini akan memberikan bilangan purata sel nanah bagi empat imej kahak ini dan menentukan skor bagi sel nanah berdasarkan nilai purata yang diperolehi. Skor ditentukan dengan merujuk kepada 'Bartlett Criteria'.

ABSTRACT

Diseases relate to lung such as *Moraxella catarrhalis*, *Mycobacterium tuberculosis* and others can be determined from sputum. However, sputum sample needs to undergo culturing process which requires high cost before the diseases can be determined. Therefore, sputum quality testing is requires to be performed on sputum sample to avoid any waste. So, only the quality or positive sample is cultured and reject the negative sample. This project is conducted to replace manual method used to determine the quality of sputum in USM, Kubang Kerian based on Modified Bartlett's Criteria. The manual method refers to the process of evaluating sputum sample by observing the sample through microscope. The technologists calculate the number of pus cells and epithelial cells through microscope to find out the score for each type of cells according to Bartlett's Criteria. So, vision system based on image processing which is able to detect and count the number of pus cells automatically is developed to enhance the manual method. This system requires at least four images of sputum from one sputum sample. The images are processed one by one through this system. The number of pus cells for each image is determined. At the end, the average number of pus cells for these four images is determined as well as its score. The score is determined by referring to the Bartlett's Criteria.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Sputum is a material coughed up from the lungs and expectorated through the mouth. Sputum contains pus cells, squamous ephitelial cells, gram-positive and gram-negative organisms. All of these elements are important to determine the quality of sputum. Sputum needs to undergo quality testing before it can be cultured to avoid high cost of culturing process if it is not a quality sputum. Culturing process is performed to detect diseases relates to lung such as *Moraxella catarrhalis*, *Mycobacterium tuberculosis* and others. Only quality sputum is allowed to undergo this process. In USM Kubang Kerian, detection and summation of these elements is manually done by human. The detection and summation is based on Modified Bartlett's criteria. Each criteria of the elements above has its own score where the total of the score will give the result of sputum quality. However, all of these elements need to be detected and counted separately. Human's eyes cannot observe through microscope for a long period of time. Thus, after 10 to 12 samples, the technologists need to rest their eyes. Therefore, automated detection and summation of pus cell for sputum quality testing system is developed to detect and count these elements automatically. However, this project just covers task to detect and count the number of pus cells in sputum image. The image was taken from x10 magnification digital biological microscope. Therefore, MATLAB R2010a is used to develop this automated detection and summation of pus cell for sputum quality testing system. Through image processing tools, some techniques are applied for detection purpose like k-means clustering, circular average filter, dilation and any other techniques which are suitable to remove

the unwanted elements in sputum image. Unwanted elements refer to squamous epithelial cells, gram-positive and gram-negative organisms. After all of these elements have been removed, the left pus cells is counted automatically.

1.1 Problem Statement

There are a lot of criteria that can be chosen to determine the quality of sputum. In HUSM, they are using Modified Bartlett's Criteria for this purpose. Based on this criteria, the number of pus cells in the image need to be determined to get the right score. Technologist in HUSM determine the number of pus cell by calculating these cells manually from sputum image. Since human might have eye drowsiness while dealing with this task, thus automated detection and summation of pus cell for sputum quality testing system needs to be developed to curb this problem. Besides that, this system can speed up and smooth the process. As a result, more than 100 samples can be processed in a day.

1.2 Objective

The objectives of this project are:

I. To develop system which is able to detect and count the number of pus cells automatically.

II. To determine the score of pus cells based on Bartlett's Criteria.

1.3 Scope of Project

This project just covers software development. Most of the techniques used are based on image processing techniques that can be applied by using MATLAB R2010a. The image taken from x10 magnification digital biological microscope is processed to get the image of pus cells from the sputum image that contains pus cells, epithelial cell, bacteria and other unwanted objects. All of the coding of this process is written on M-File command window whereas the results are shown through the final figure. The number of pus cells is also shown in final figure. However, Graphical User Interface (GUI) is developed to present this project properly. Thus, the focuses of this project are:

I. Detect and count the number of pus cells in sputum image.

II. Determine the score of pus cell in sputum sample based on Modified Bartlett's Criteria.

CHAPTER II

LITERATURE REVIEW

Based on *A.S.Abdul Nasir* writing, he states that leukaemia is a blood cancer that causes more deaths than any others cancer among children and young adults under the age of 20. There are two major types of acute luekaemia such as Acute Lymphobastic Leukaemia (ALL) and Acute Myelogenous Leukaemia (AML). In leukaemia diagnosis, size and shape of abnormal white blood cell (WBC) would be observed by haematologists in order to differentiate the types of acute leukaemia.

For the morphological analysis of acute leukaemia images, *A.S.Abdul Nasir* proposed the combination between linear contrast technique and colour segmentation based on HSI (Hue, Saturation, Intensity). This type of colour space was used in order to obtain a fully segmented abnormal WBC and nucleus of acute leukaemia images. Then, k-means clustering algorithm is used to ease the segmentation process. Next, the fully segmented WBC which consists of cytoplasm and nucleus regions can be achieved by using the combination of linear contrast technique and segmentation based on H component image. Meanwhile, the fully segmented nucleus can be obtained by applying the segmentation based on S component image.

These techniques have produced a better effect on improving the accuracy of WBC segmentation with segmentation accuracies of 99.02% and 99.05% for segmented WBC and nucleus, respectively.

2.1 Sputum Quality Testing

Previous study has shown that the implementation of rejection criteria in clinical specimens saves the reagent cost up to US$28000 and 1082 hours of technologist time [1]. It means that culturing process requires high cost and takes time to be performed. Thus, sputum should undergo sputum quality testing before it can be cultured. Then, only sputum with quality is allowed to be cultured. It is done to reduce high cost of culturing process and save technologist time because the culturing process on non quality sputum will waste the cost and time due to zero output. There are six different criteria for judging the acceptability of sputum specimens [11]. All of the criteria, method and criteria for acceptability are summarized in the table1.

Table 2.1: Summary of six published criteria for judging acceptability of sputum specimens.

Criteria	Method	Criteria for acceptability
Bartlett	Assign + and – values: +2 if >25 Neutrophils;+1 if 10-25 Neutrophils; +1 if mucus seen; -2 if >25 EPI;-1 if 10-25 EPI	Any positive score (sum of + and – values assigned)
Murray and Washington	Average no. of EPI/LPF	<10 EPI/LPF
Geckler et al.	Average no. of EPI/LPF	<25 EPI/LPF
Van Scoy	Average no. of Neutrophils/LPF	>25 Neutrophils/LPF
Barry	Assign + and - values: +3 if >150 Neutrophils; +2 if 76-150 Neutrophils; +1 if 1-75 Neutrophils; -3 if >25 EPI; -2 if 16-25 EPI; -1 if 5-15 EPI	Any positive score (sum of + and – values assigned)
Heineman and Radano	Average ratio of Neutrophils to EPI	>10 Neutrophils/EPI

In which,

Neutrophils= pus cells

EPI=squamous epithelial cell

LPF = Low Power Field

2.2 Modified Bartlett's Criteria

The quality of sputum is determined by using Modified Bartlett's criteria in USM Kubang Kerian. Since this project has collaboration with USM Kubang Kerian, thus Modified Bartlett's Criteria is also applied to develop this automatic vision system. Modified Bartlett's Criteria is chosen due to its easiness in interpretation and lower rejection rate than other criteria, thus minimizing the number of missed potential pathogen [1]. Bartlett's criteria is based on the relative number of inflammatory cells, squamous epithelial cells and mucus seen in Gram-stained smears. However, there are some differences in Modified Bartlett's Criteria, whereby the macroscopic appearances of the sample were taken into account:wether they are mucoid, mucopurelent, purelent or blood stained [1]. This criteria uses total score of each criteria to determine the quality of sputum. The score of each criteria is summarized in the Table 2.2.

Table 2.2: Modified Bartlett's Criteria

	Criteria	Score
Neutrophils (pus cells) count (Score A)	< 10 neutrophil/10x field	0
	10-25 neutrophils/10x field	+1
	>25 neutrophils/10x field	+2
Macroscopy (Score B)	Mucoid, Mucopurelent, Purelent, or Blood stained	+1
Squamous epithelial cell count (Score C)	< 10 Squamous epithelial cell/10x field	0
	10-25 Squamous epithelial cell /10x field	-1
	>25 Squamous epithelial cell /10x field	-2

The decision either sputum sample can be cultured or not is depends on the total score. Total score refers to the summation of score A, score B and score C (Total score = scoreA + scoreB +scoreC). The sputum will be cultured if total score is 1 and above. The sputum is rejected for the total score of 0 and below which can be classified as non quality sputum.

2.3 Pus Cells

Eventhough sputum contains pus cells, squamous epithelial cells and other bacteria but this project just focus on detection and summation of pus cell in sputum image. Thus, most of the explanation will relate to pus cells. Pus cell or Polymorphonuclear Neutrophils (PMNs) play a central role in innate immunity, where they dominate the response to infections, in particular in the cystic fibrosis lung. PMNs are phagocytic cells that produce a wide range of antimicrobial agents aimed at killing invading bacteria [12]. PMNs migrate from the blood stream to the injured tissues where they eliminate invading organisms by phagocytosis and killing [5]. The size of pus cell is

smaller than squamous epithelial cell [6]. The example of pus cell, squamous epithelial cell and bacteria in sputum image are shown in Figure 2.1.

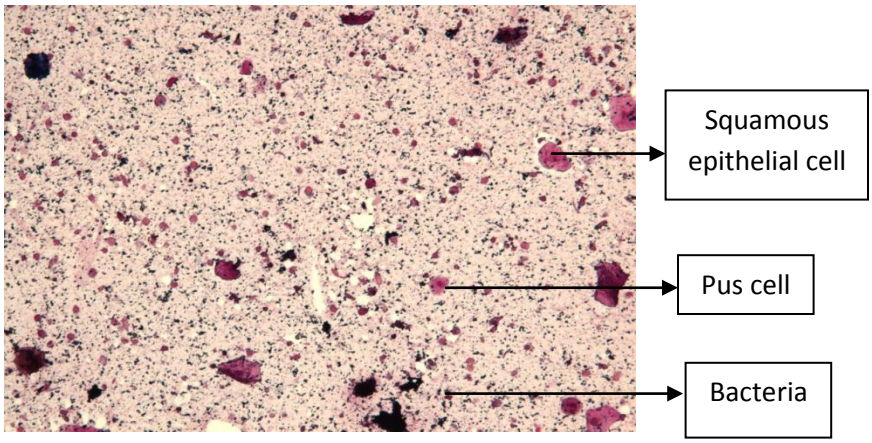


Figure 2.1 Sputum image under x10 computerized microscope.

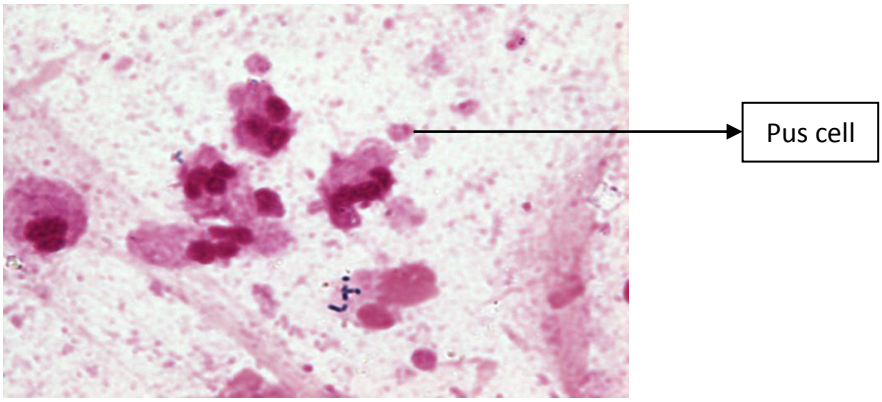
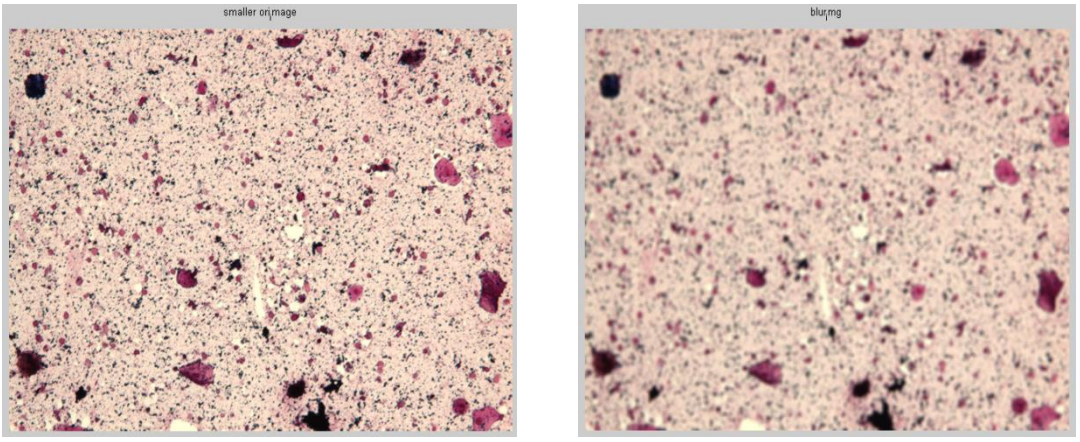


Figure 2.2 Pus cell image under x100 computerized microscope.

2.4 Image Enhancement: Circular Average Filtering

Out of focus blur is usually modeled as circular averaging filter (pillbox) [7]. In MATLAB, this filtering process is done by the command fspecial('disk', radius) that

returns the circular averaging filter (pillbox) within the square matrix of side 2*radius+1. The default radius is 5 [8]. This command softens the edges of the function, alleviating the discrete nature the circular would otherwise demonstrate. Discretely defined circles tend to have edges that are jagged. These jagged edges introduce frequency components that would not be present in a real circular aperture. These frequencies act as noise in the pattern and corrupt the frequency space diagram. To avoid this problem, the 'soft' edged circular should be used. The example of circular averaging filter performed on image is shown in figure 2.



(a) (b)

Figure 2.3 (a) Original sputum image, (b) Image after circular average filtering with radius 3 is performed.

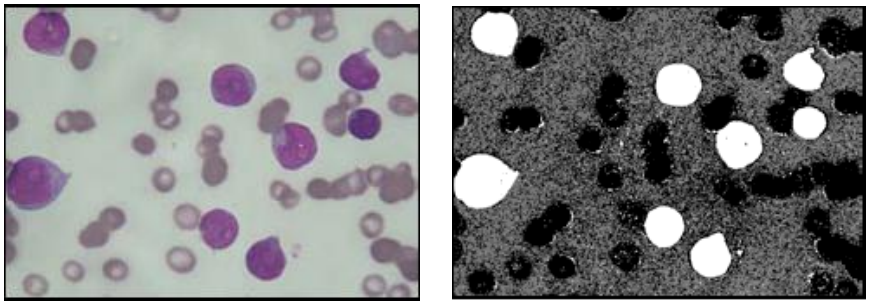
2.5 Colour Segmentation: K-Means Clustering Algorithm

K-Means algorithm classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them [2].

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (2.1)$$

Where there are k clusters $s_{i,,}$ $i = 1,2,\ldots,k$ *and* μ_i is the centroid or mean point of all the points $x_j \in s_i$

The cluster should exhibit two properties, they are (1) each group must contain at least one object (2) each object must belong to exactly one group [3]. K-means clustering is used because it is simple and has relatively low computational complexity [4]. In addition, it is suitable for biomedical image segmentation as the number of clusters (K) is usually known for images of particular regions of human anatomy [4]. For smaller values of k the k-Means algorithms give good results [2]. For larger values of k, the segmentation is very coarse, many clusters appear in the images at discrete places [2]. This is because Euclidean distance is not a very good metric for segmentation processes [2]. The example of k-Means clustering is shown in Figure 2.4 with k=3.



(a) (b)

Figure 2.4 (a) original image (b) image after k- Means clustering (three clusters)

2.6 Morphological Operation: Dilation

Dilation 'grows' or 'thickens' objects in a binary image. The specific manner and extent of this thickening is controlled by the shape of the structuring element used [9]. The dilation of A by B, denoted $A \oplus B$ is defined as

$$A \oplus B = \{z|(B)_z \cap A \neq \phi\} \quad (2.2)$$

Where Φ is empty set and B is structure element [10].Dilation operation is shown [10] in figure 2.5.

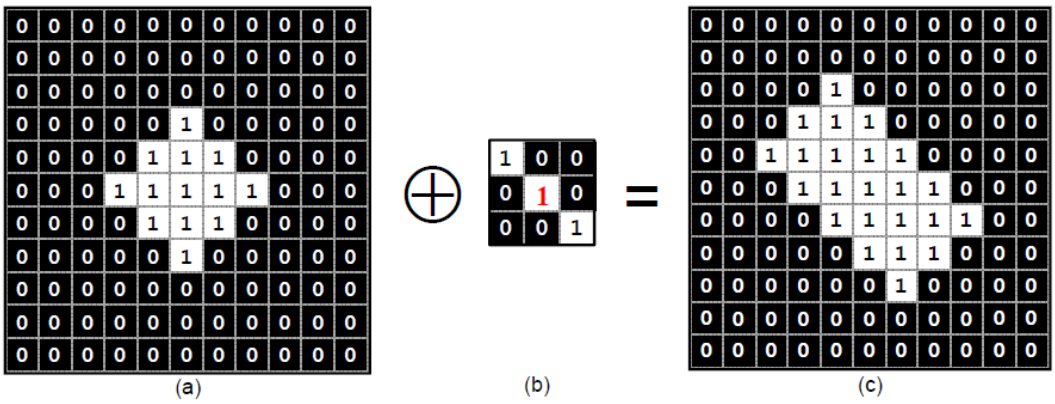


Figure 2.5 Illustration of morphological operation. (a) Original binary image with diamond object, (b) Structure element with three pixels arranged in a diagonal line, the origin of structure is identified by red 1, (c) Dilated image.

Example of dilation operation [10]:

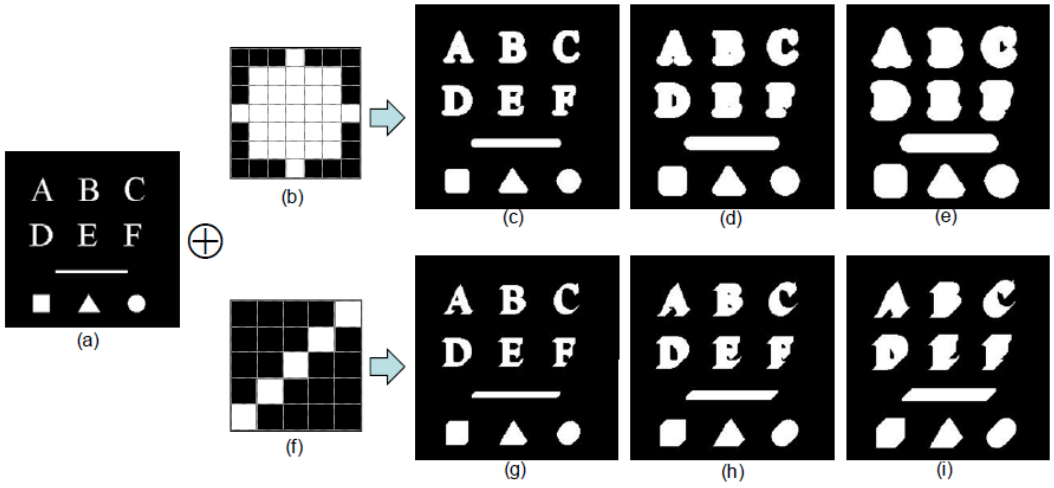


Figure 2.6 Example of morphological operation, (a) binary input image, (b) a disk structure element, (c) Dilated image of (a) by SE (b), (d) after twice dilation of SE (b), (e) after three times dilation of SE (b), (f) a line structure element, (g) dilated image of (a) by SE (f), (h) after twice dilation by SE (f), (i) after three times dilation by SE (f).

2.7 Edge Detection

Edge detection is the approach for segmenting image based on local changes in intensity. It is the most common approach for detecting meaningful discontinuities in gray level [13]. Such continuities are detected by using first and second-order derivatives. The first order derivatives of choice in image processing is the gradient of a 2-D function, *f(x,y),* is defined as the vector [13]

$$\nabla f \equiv grad(f) \equiv \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix} \tag{2.3}$$

The general syntax for this function is [g,t] = edge(f, 'method', parameters) [13]

In MATLAB R2010a, there are some types of edge detector that can be used to detect edge in image. The summary of these edge detectors is listed in the table 2.3 [13]